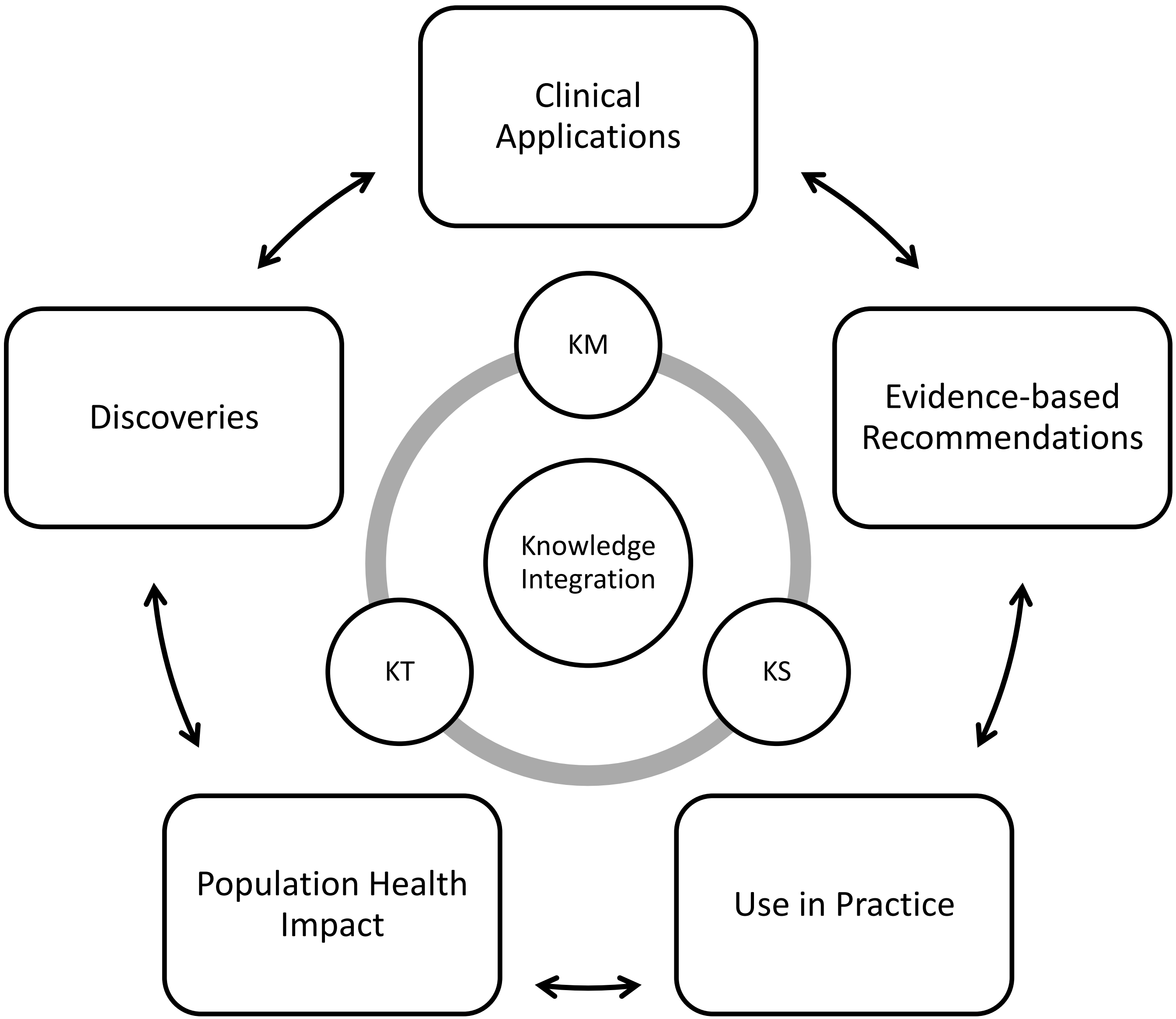
The role of epidemiology in knowledge integration and meta-research

John P.A. Ioannidis

C.F. Rehnborg Professor in Disease Prevention
Professor of Medicine and Professor of Health Research and Policy
Director, Stanford Prevention Research Center
Stanford University School of Medicine
Professor of Statistics (by courtesy)
Stanford University School of Humanities and Sciences



Adapted from *Cancer Epidemiol Biomarkers Prev*. 2013; 22(1):3-10, Knowledge Integration in Cancer: Current Landscape and Future Prospects, with permission from AACR.

Table 1. Published articles in cancer literature

	PubMed	Treatment	Prevention	Not (trial or treatment)
Cancer	2,673,926	1,360,697	208,187	1,295,958
+Animal	481,080	206,009	40,891	269,653
+Cell	1,118,600	510,286	66,935	604,187
+Cohort	53,567	29,808	8,683	22,819
+Case–control	59,248	11,255	5,848	26,973
+Risk	267,490	158,529	58,002	105,276
+Biomarker	204,419	91,298	15,675	111,025
+Clinical trial, type	105,939	93,172	14,807	4,535
+RCT, type	34,449	31,862	8,047	0
+Meta-analysis, type	6,406	3,902	1,153	2,332
+Systematic review, type	28,922	21,398	5,755	6,763
+Review, type	314,176	201,162	39,859	111,234

NOTE: Search strategies: treatment: "treatment", prevention: "prevention OR screening [ti] OR screening [tw]", Not trial or treatment: "NOT (trial* OR treatment)."

Abbreviation: RCT, randomized controlled trial.

Table 2. Different approaches to KM and KS

Knowledge management	
	Published data: optimization of search engines, curation and cleaning, harmonization
	Unpublished data: detection, registration, cleaning
	Deposition of raw datasets in public: documentation, access control, ease of use, credit, independence
	Live stream information
Knowledge synthesis	
	Same-level of information
	Systematic reviews of published information
	Meta-analyses of published information
	Meta-analyses including also retrieved unpublished data
	Field synopses of many meta-analyses
	Collaborative meta-analyses of previously collected individual-level data
	Collaborative meta-analyses of prospectively collected data from existing studies
	Prospective consortia and meta-analyses thereof
Multiple levels of information	
	Cross-design synthesis and multilevel evidence appraisals
	Modeling with real or simulated data
	Meta-research (research on research)

Table 3. Systematic reviews and meta-analyses in different fields of cancer (excluding trials and treatment)

Search terms	All articles, *N* (all)	*n* (SR)	*n* (MA)	*n* (SR)/*n* (MA)	*n* (all)/*n* (MA)
Gene/genome/genetic	268,597	1,999	920	2.2	291
Epigenetic/methylation/mutation	115,763	497	159	3.1	728
Immune/allergy/asthma	29,046	107	25	4.3	1162
Hormone	53,679	148	51	2.9	1,032
Social/socioeconomic	11,531	224	50	4.5	231
Diet/dietary/nutrition/nutritional	19,549	289	147	2.0	133
Physical activity/exercise/obesity	8,919	192	74	2.6	121
Virus/bacteria/infection/infectious	88,881	331	109	3.0	815
Carcinogen	30,286	201	88	2.3	344
Radiation	30,124	229	86	2.6	350
Occupation/occupational	16,839	344	177	1.9	95
Smoking/smoke/tobacco	20,660	413	232	1.8	89
Alcohol	17,921	170	83	2.0	215
Biomarker	72,709	373	97	3.8	750

Examples of knowledge integration at the meta-research level

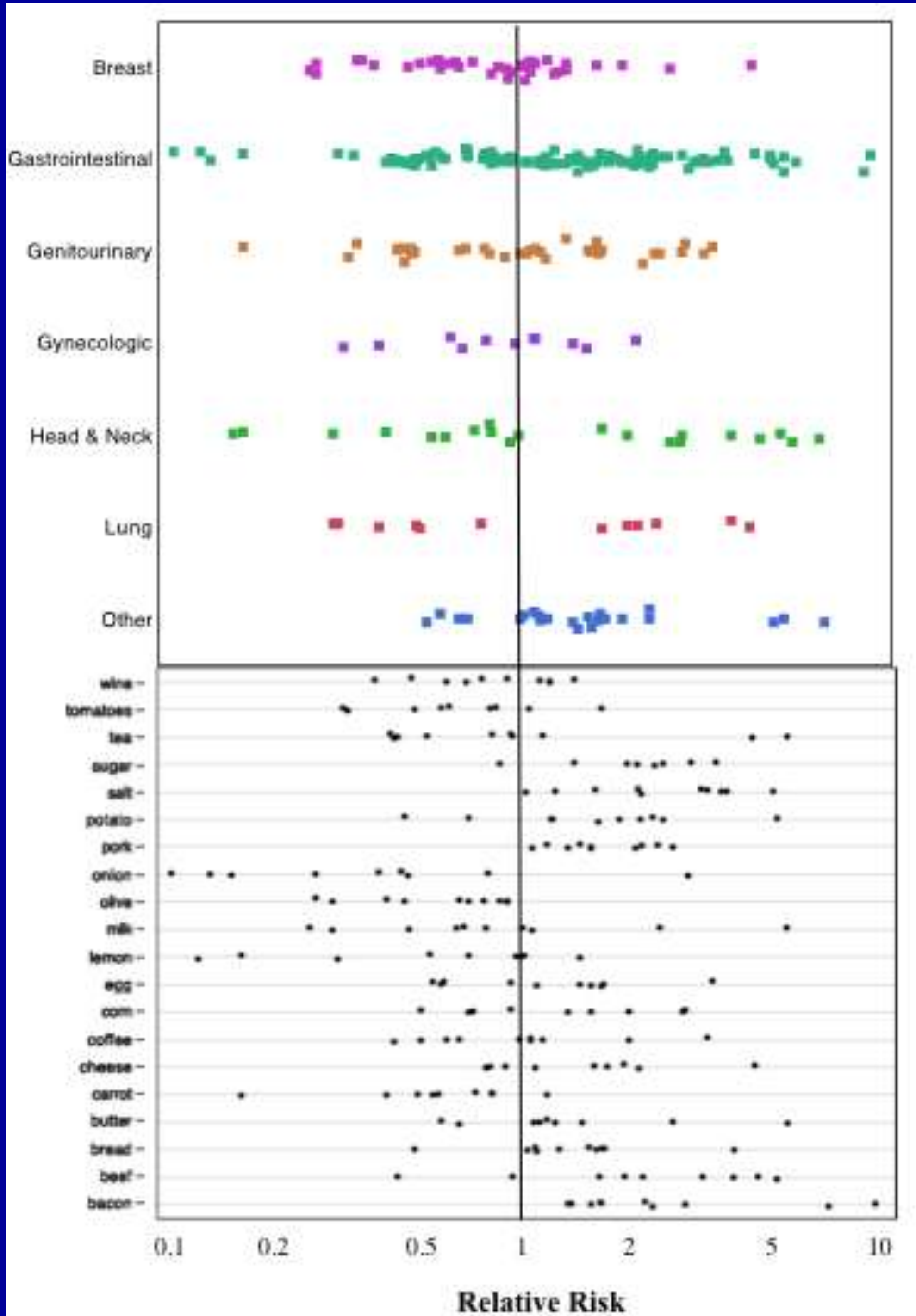
- Associations
- Predictions
- Treatments

Nutrients and cancer risk

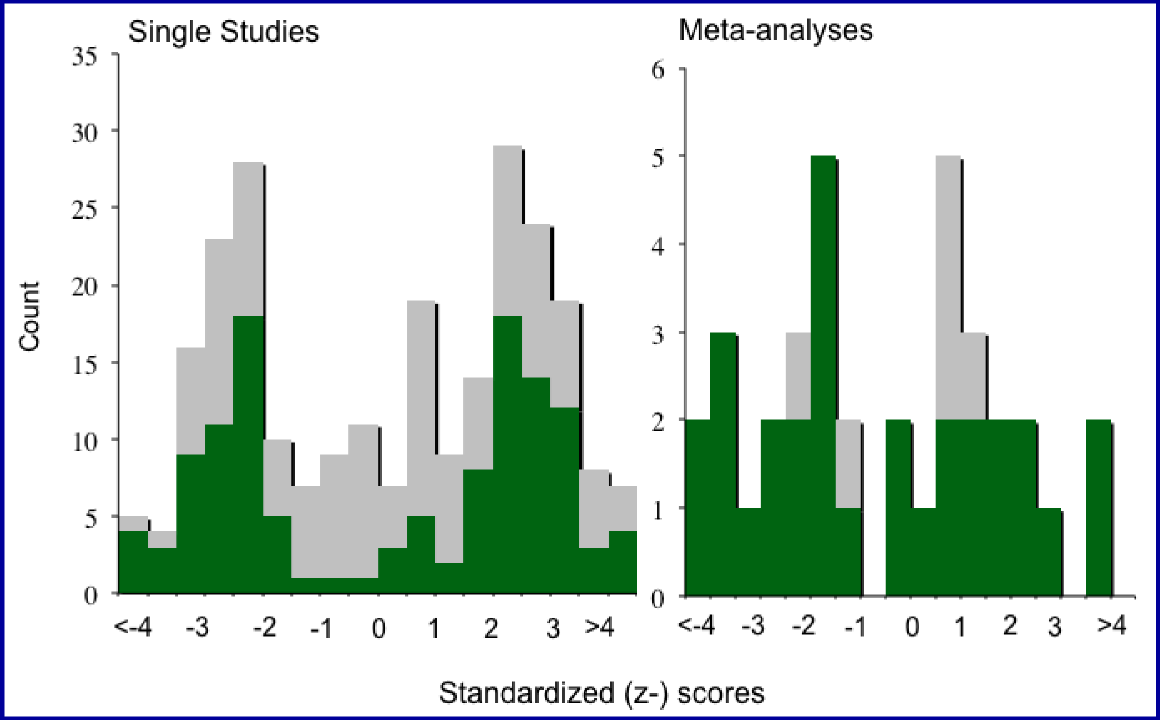
- Open a popular cookbook
- Randomly check 50 ingredients
- How many of those are associated with significantly increased or significantly decreased cancer risk in the scientific literature?

Associated with cancer risk

- veal, salt, pepper spice, flour, egg, bread, pork, butter, tomato, lemon, duck, onion, celery, carrot, parsley, mace, sherry, olive, mushroom, tripe, milk, cheese, coffee, bacon, sugar, lobster, potato, beef, lamb, mustard, nuts, wine, peas, corn, cinnamon, cayenne, orange, tea, rum, raisin

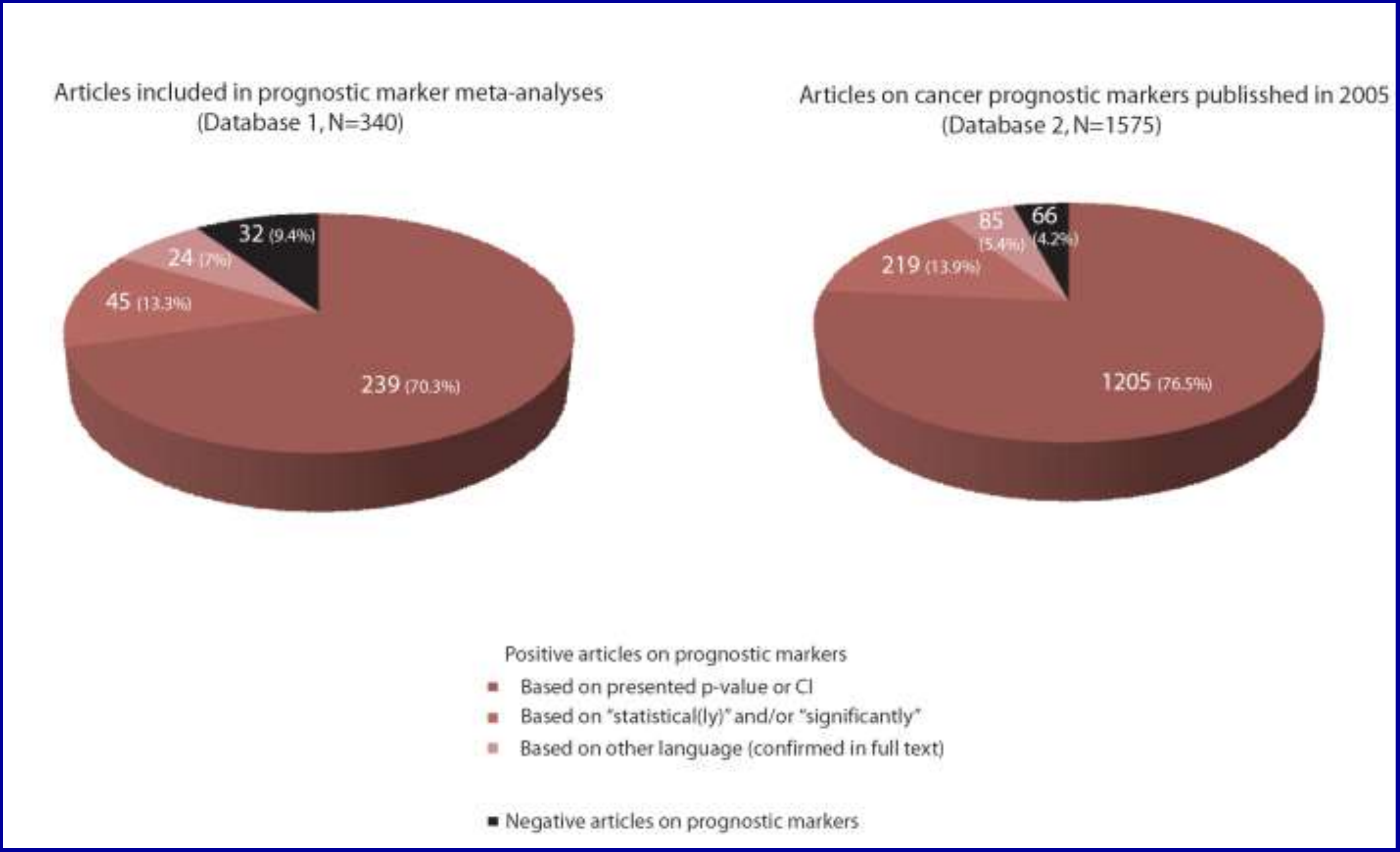


Reprinted from Schoenfeld JD and Ioannidis JPA. Is everything we eat associated with cancer? A systematic cookbook review. *Am J Clin Nutr.* 2013; 97(1):127-134, American Society for Nutrition.



Reprinted from Schoenfeld JD and Ioannidis JPA. Is everything we eat associated with cancer? A systematic cookbook review. *Am J Clin Nutr.* 2013; 97(1):127-134, American Society for Nutrition.

Prognostic tumor markers



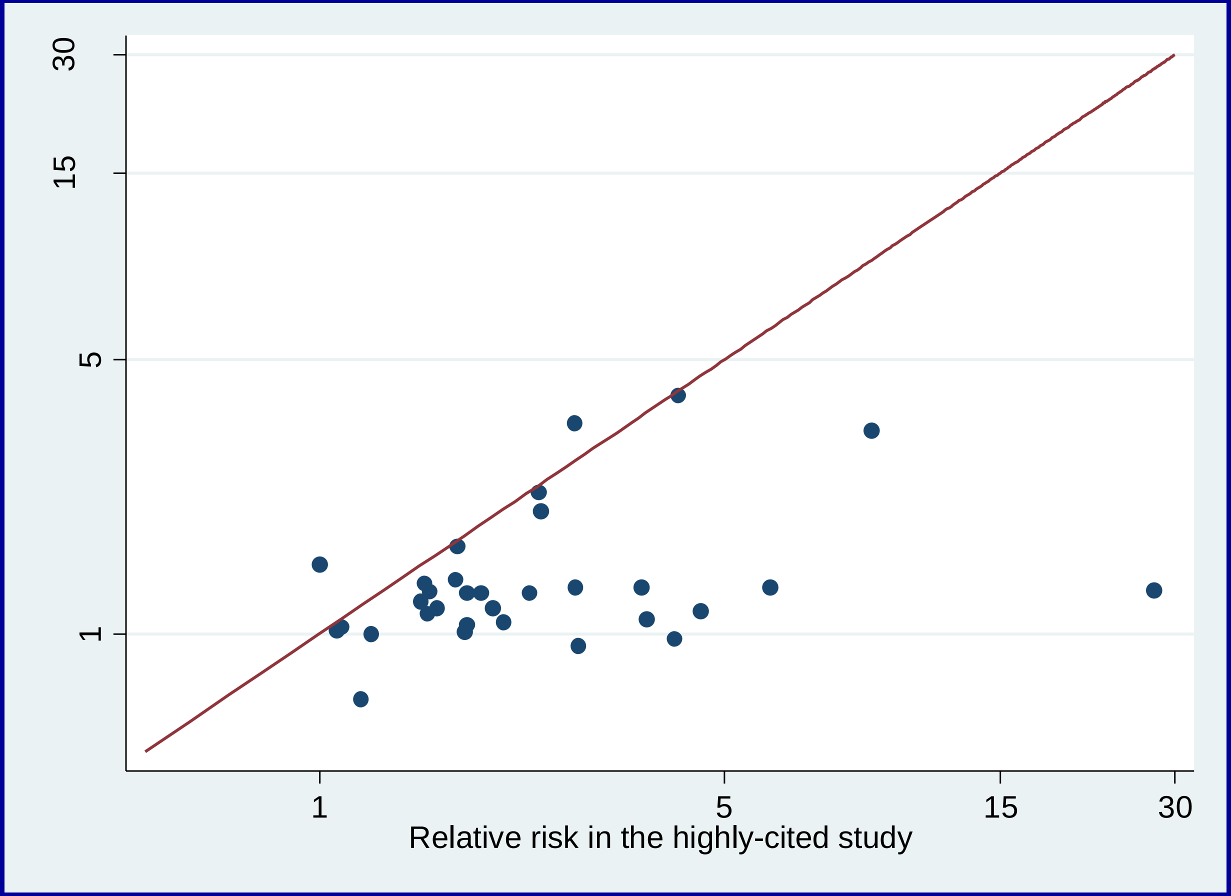
Adapted from Kyzas PA, Denaxa-Kyza D, and Ioannidis JPA. *Eur J Cancer*. 2007; 43(7): 2559-2578.

Further analysis of claims in “negative” prognostic studies

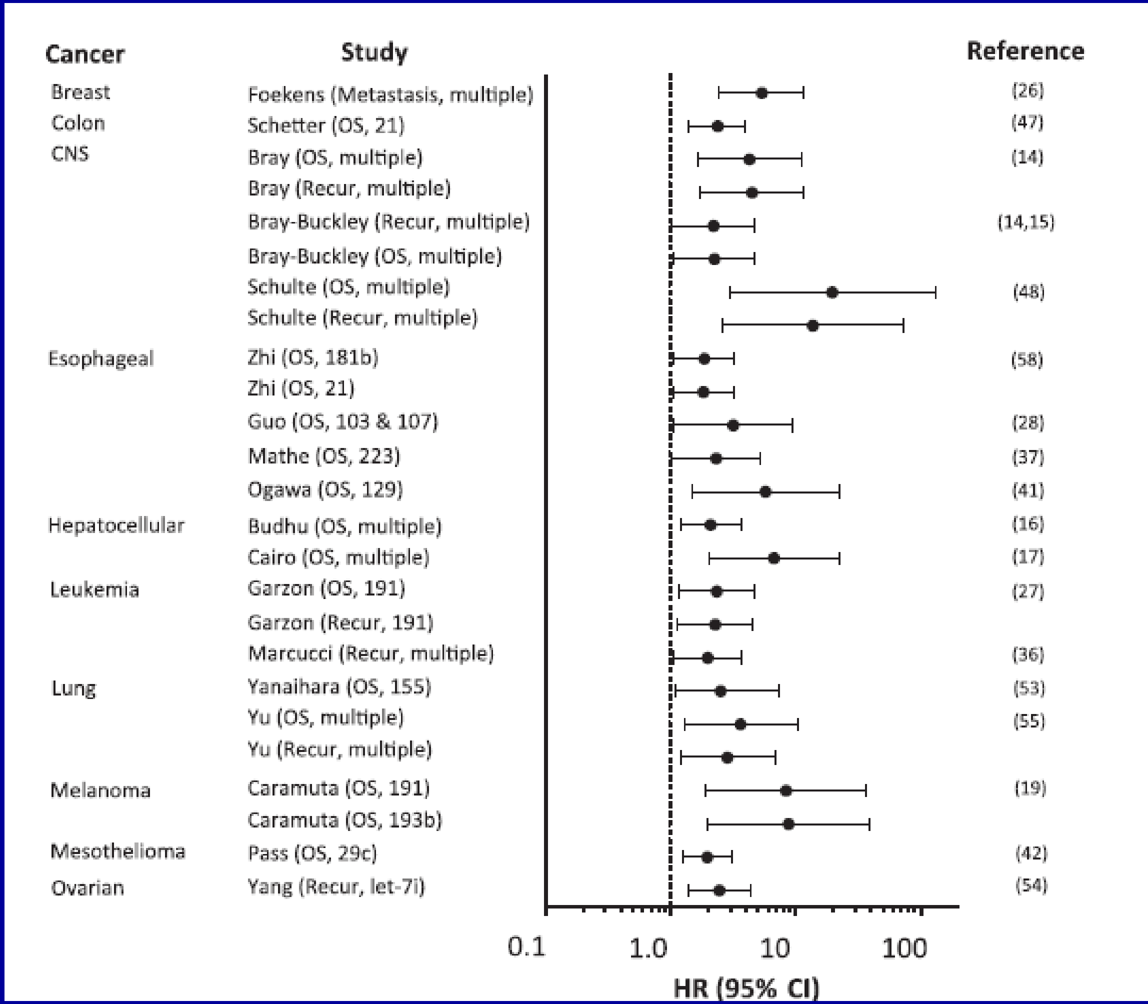
	Database 1 N (%)	**Database 2** N (%)
Not admitted to be fully “negative”	27 (7.9)	45 (2.8)
Significance for other (non-prognostic) analyses	6 (1.7)	11 (0.6)
Discussion of non-significant trends	2 (0.6)	5 (0.3)
Offered apologies	9 (2.8)	13 (0.8)
Significance for other analyses + Discussion of non-significant trends	1 (0.3)	3 (0.2)
Significance for other analyses + Offered apologies	6 (1.7)	7 (0.5)
Discussion of non-significant trends + Offered apologies	3 (0.8)	4 (0.3)
All three mechanisms	-	2 (0.1)
Admitted to be fully “negative”	5 (1.5)	21 (1.3)

Adapted from Kyzas PA, Denaxa-Kyza D, and Ioannidis JPA. *Eur J Cancer*. 2007; 43(7): 2559-2578.

Effect sizes for the top-cited biomarkers in the biomedical literature



Adapted from Ioannidis JPA, Panaigotou OA. *JAMA*. 2011; 305(21):2200-2210.

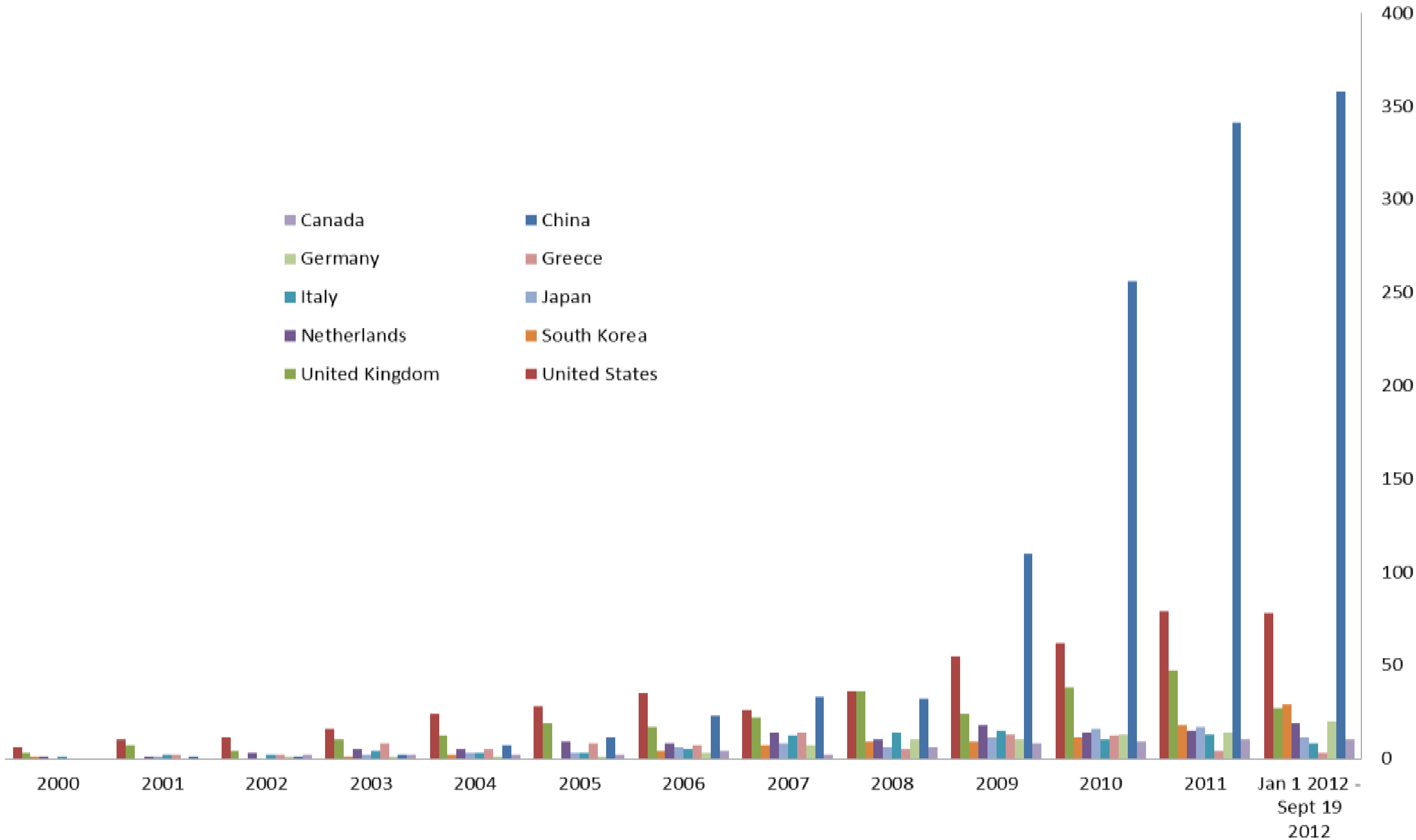


Nair VS, Maeda LS, and Ioannidis JPA. Clinical outcome prediction by microRNAs in Human Cancer: A systematic review. *J Natl Cancer Inst*. 2012; 104(7): 528-540. Used by permission of Oxford University Press.

Field synopses

Chatzinasiou F, Lill CM, Kypreou K, et al. Comprehensive Field Synopsis and Systematic Meta-analyses of Genetic Association Studies in Cutaneous Melanoma. *J Natl Cancer Inst*. 2011; 103(16): 1227-1235.

Published Genetic Meta-Analyses by Country, Year



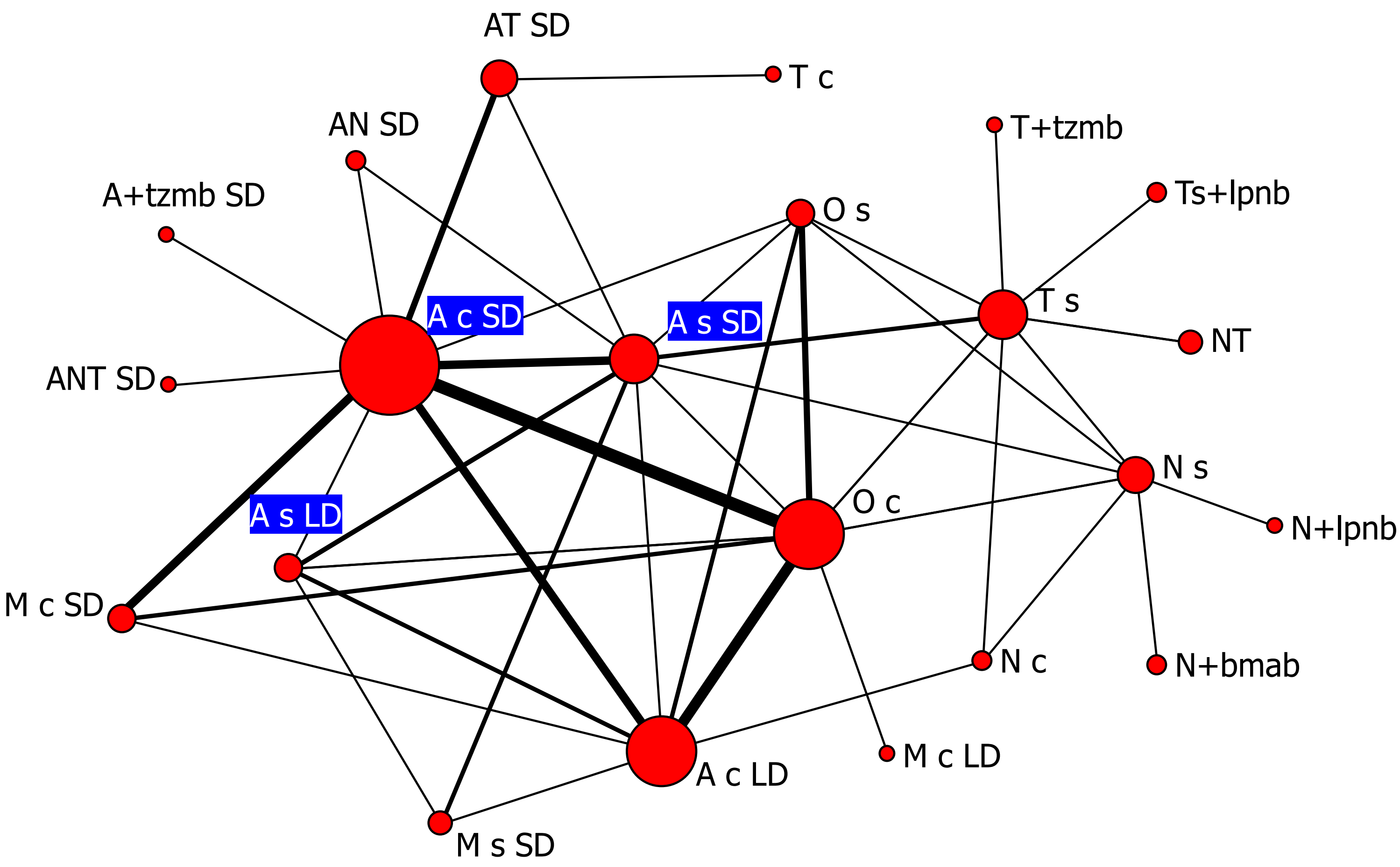
Ioannidis JPA, Chang C, Lam TK, Schully SD, and Khoury MJ. Submitted to *Cancer Epidemiol Biomarkers Prev*, 2013.

Replicated: only 6 of 53 landmark studies for Amgen oncology drug target projects

- "The failure to win "the war on cancer" has been blamed on many factors, … But recently a new culprit has emerged: too many basic scientific discoveries… are wrong."

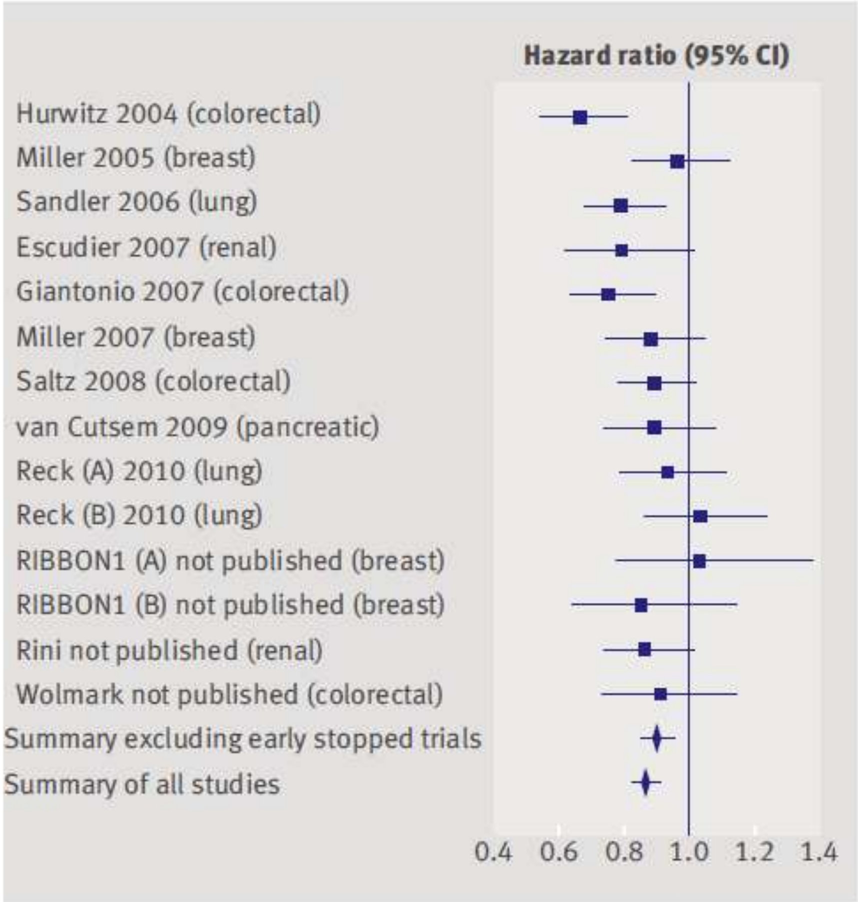
Source: Begley S. *Reuters*. March 25, 2012.

700 randomized trials on advanced breast cancer: it all works (more or less)



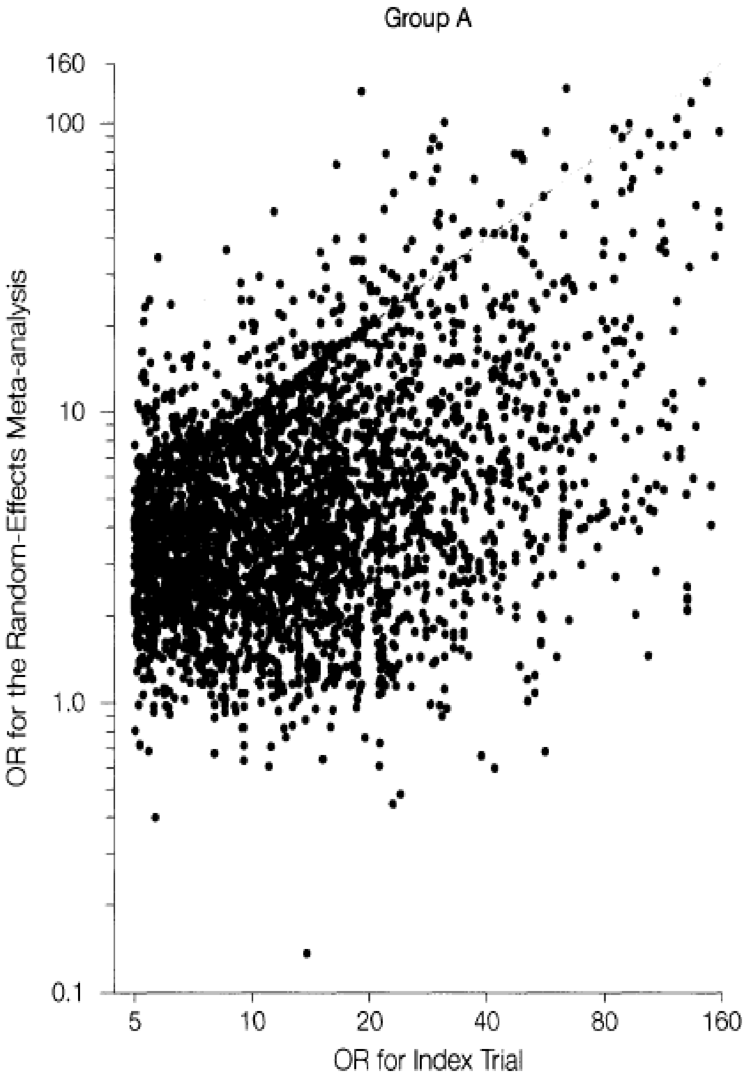
Adapted from Mauri D, Polyzos NP, Salanti G, et al. *J Natl Cancer Inst*. 2008; 100(24):1780-1791.

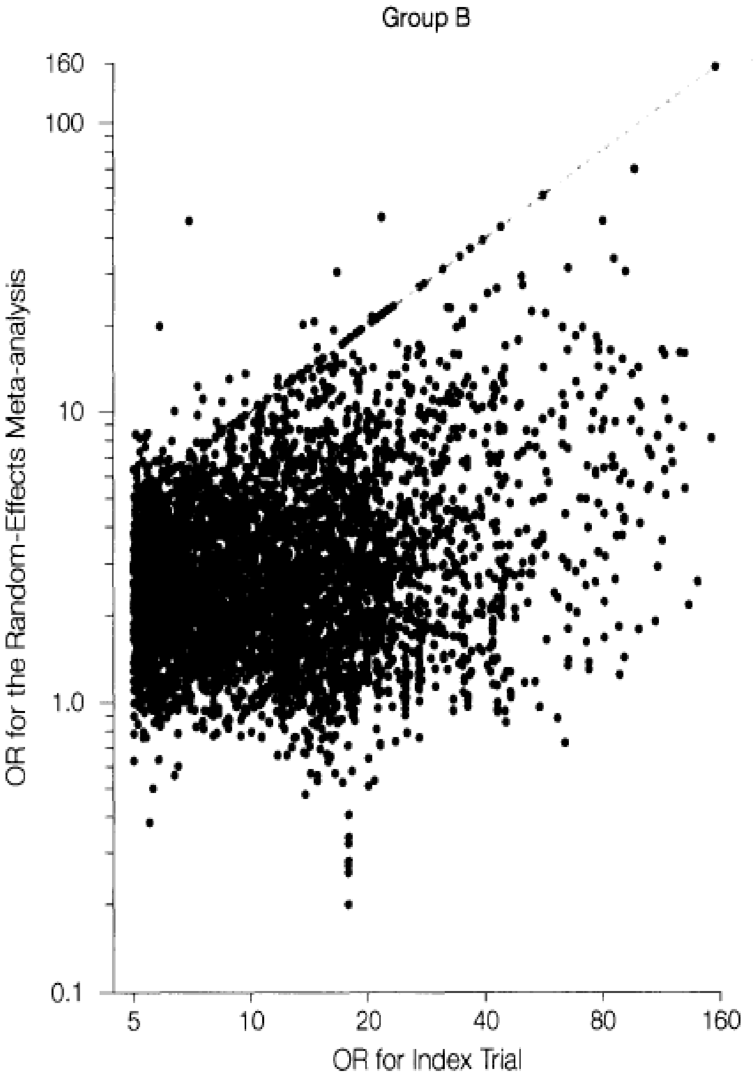
1200 (and counting) clinical trials of bevacizumab



Meta-analysis forest plot for survival with bevacizumab *v* control in trials of patients with cancer. Each trial is shown by its year of publication, name of first author, and type of malignancy as well as the hazard ratio for survival and 95% confidence interval. Also shown are summary estimates including all trials and excluding the three trials stopped early, which showed large treatment benefits (Hurwitz 2004, Sandler 2006, Escudier 2007)

Pereira TV, Horwitz RI, and Ioannidis JPA. Empirical Evaluation of Very Large Treatment Effects of Medical Interventions. *JAMA.* 2012; 308(16):1676-1684.



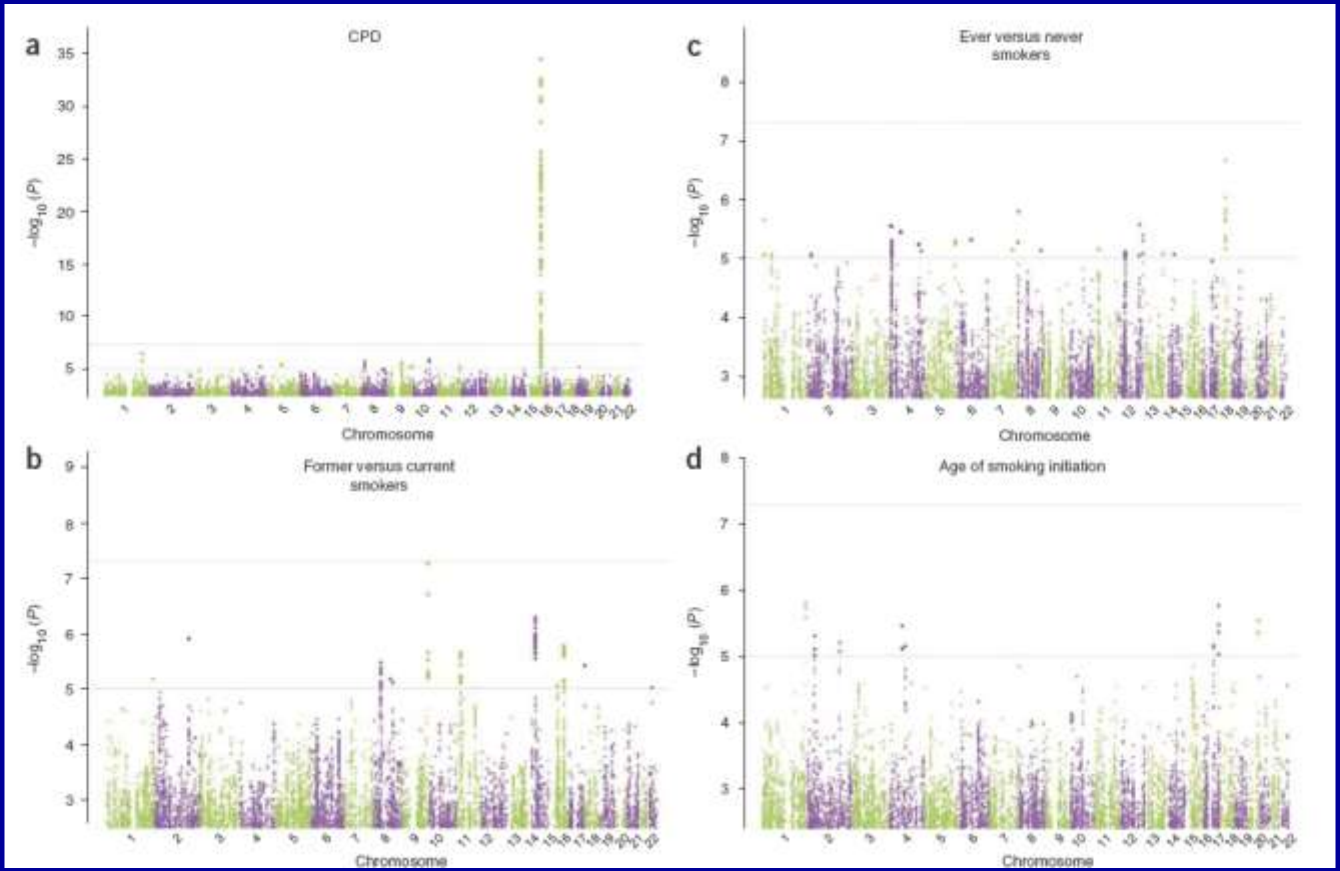


Pereira TV, Horwitz RI, and Ioannidis JPA. *JAMA.* 2012; 308(16):1676-1684.

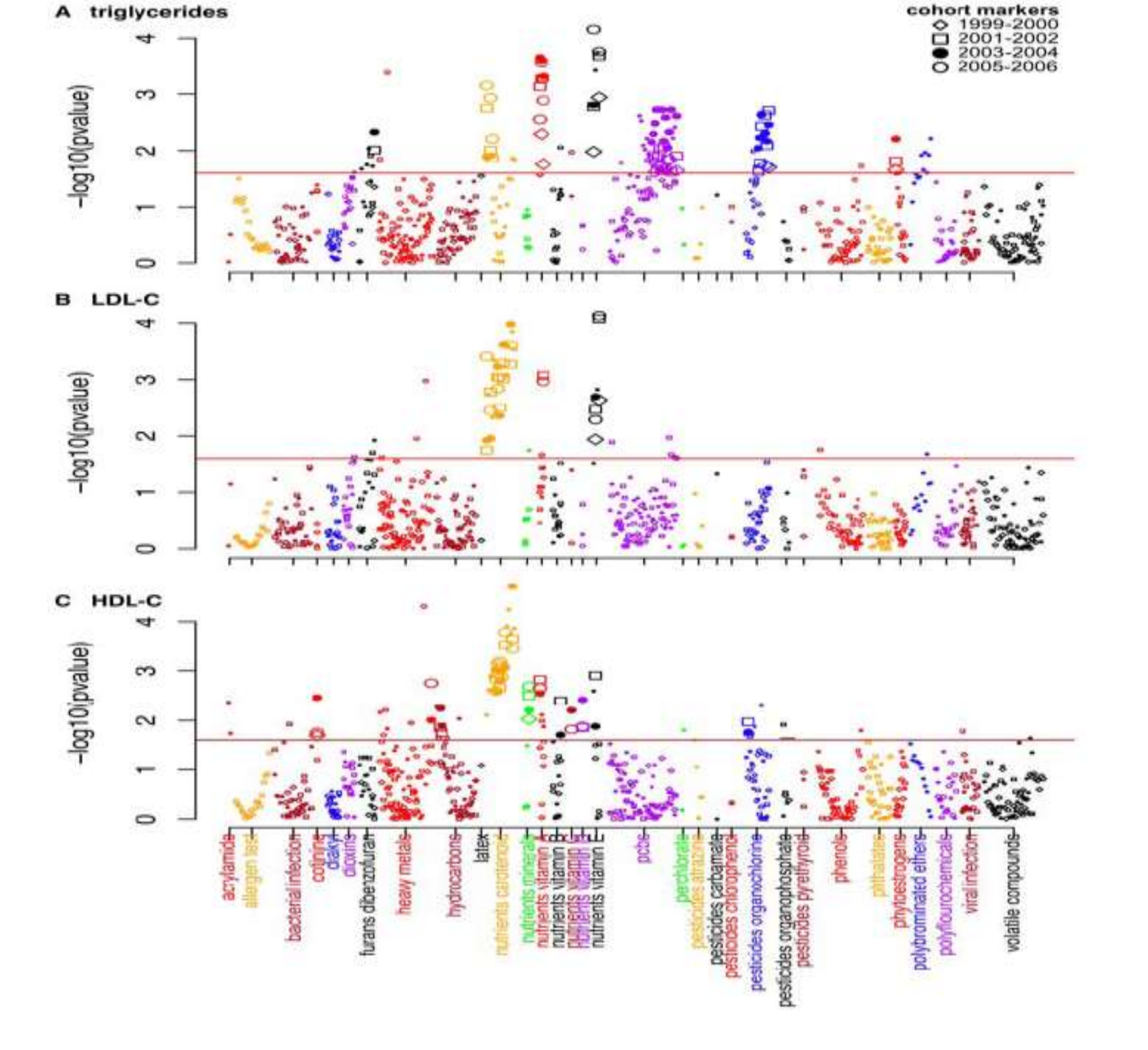
Learning to live with small/tiny effects

Siontis GCM and Ioannidis JPA. Risk factors and interventions with statistically significant tiny effects. *Int J Epidemiol*. 2011; 40(5):1292-1307.

Large-scale collaboration

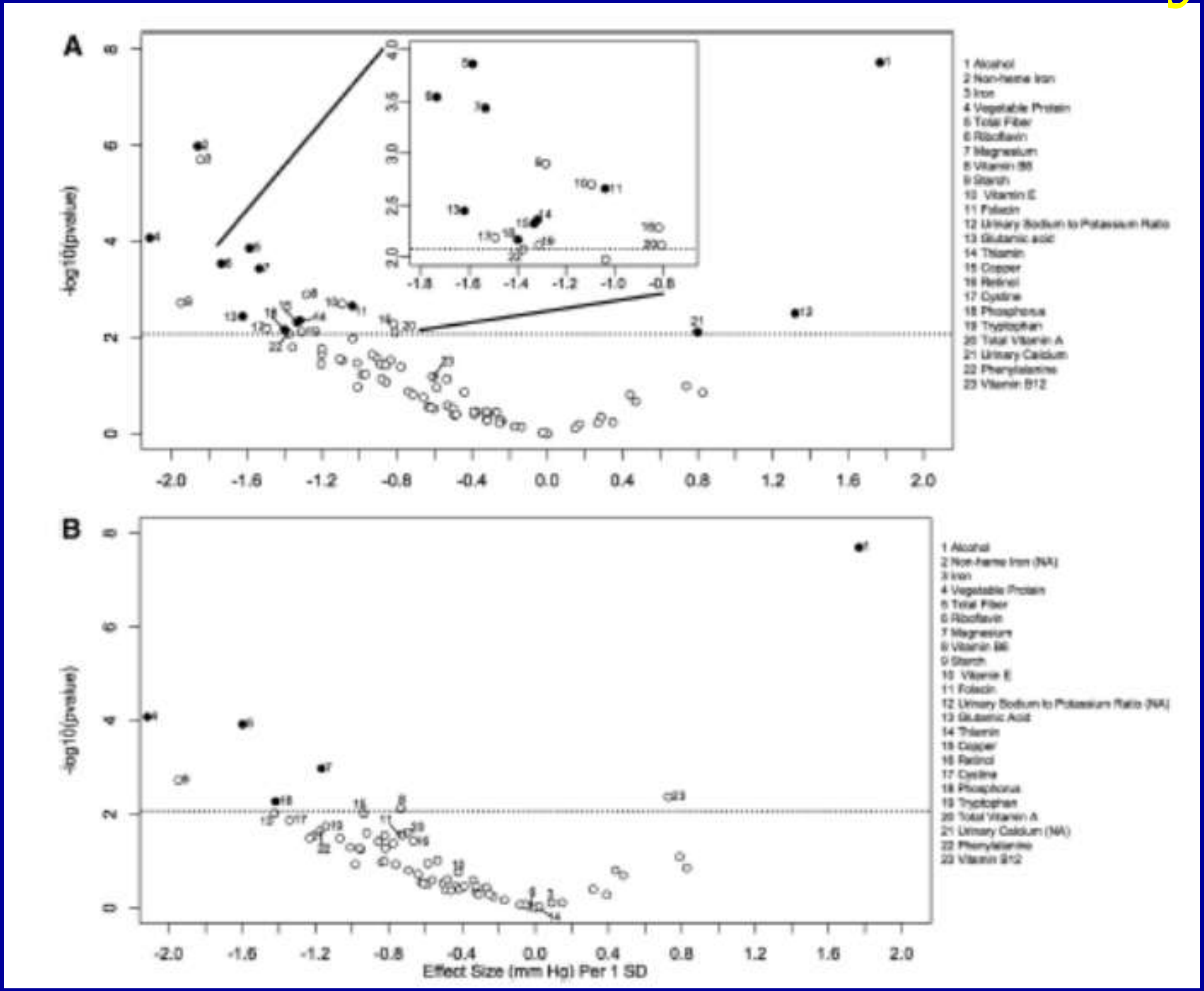


Taking multiplicity into account



Adapted from Patel CJ, Cullen MR, Ioannidis JP, et al. *Int J Epidemniol.* 2012; 41(3):828-43.

Nutrient-wide association study



Reprinted from Tzoulaki I, Patel CJ, Okamura T, et al. *Circulation.* 2012; 126(21):2456-2464. Reprinted with permission from Lippincott, Williams, and Wilkins, 2012.

Validated heatmaps for nutrients

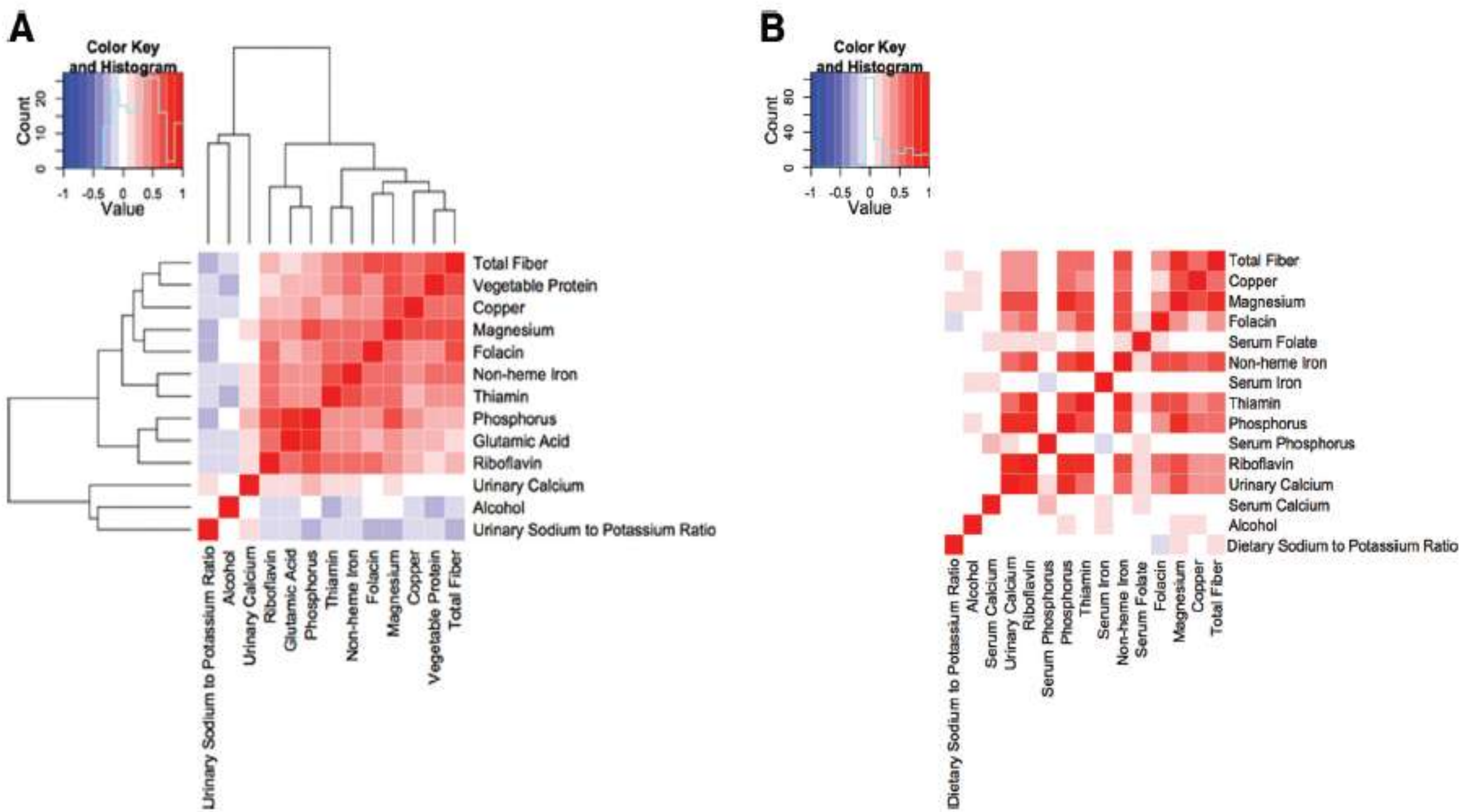


Figure 4. Pearson coefficient correlation heatmap showing all nutrients and potential confounders examined in (**A**) INTERMAP total population and (**B**) NHANES total population. Nutrients are clustered according to a hierarchical clustering algorithm in INTERMAP, grouping highly correlated factors closer to one another. For NHANES, the clustering of INTERMAP samples has been used. Correlation coefficients are adjusted for age, sex, and sample (INTERMAP)/cohort (NHANES). INTERMAP indicates International Collaborative Study on Macro-/Micronutrients and Blood Pressure; NHANES, National Health and Nutrition Examination Survey.

Improving research reporting standards: STROBE-ME, 2011

Gallo V, Egger M, McCormack V, et al. STrengthening the Reporting of OBservational studies in Epidemiology – Molecular Epidemiology (STROBE-ME): An Extension of the STROBE Statement. *PLoS Med*. 2011; 8(10).

Improving research reporting standards: GRIPS, 2011

Janssens AC, Ioannidis JPA, van Duijn CM, et al. Strengthening the reporting of Genetic Risk Prediction Studies: the GRIPS statement. *PLoS Med*. 2011; 8(3).

Registration

Ioannidis JPA. The Importance of Potential Studies That Have Not Existed and Registration of Observational Data Sets. *JAMA*. 2012; 308(6): 575-576.

Levels of registration

- Level 0: no registration
- Level 1: registration of dataset
- Level 2: registration of protocol
- Level 3: registration of analysis plan
- Level 4: registration of analysis plan and raw data
- Level 5: open live streaming

Alsheikh-Ali AA, Qureshi W, Al-Mallah MH, et al. Public Availability of Published Research Data in High-Impact Journals. *PLoS One*. 2011; 6(9).

		Policy of Required Public Deposition for Types of Data				Policy of Provision of Materials and Methods			Full data deposited
Journal	Impact Factor	Microarray	Nucleic Acid	Protein	Macromolecular	Materials upon request	Protocols upon request	Condition of publication	Percentage of papers
New England Journal of Medicine	52.589								0
Cell	29.887								1
Nature	28.751								0
Lancet	28.638								0
Nature Medicine	26.382								0
Science	26.372								1
Nature Immunology	26.218								9
Nature Genetics	25.556								0
JAMA	25.547								1
Nature Biotechonology	22.848								5
Nature Materials	19.782								0
Immunity	19.266								0
Nature Cell Biology	17.623								0
Journal of Clinical Investigation	16.915								0
Archives of General Psychiatry	15.976								0
Journal of the National Cancer Institute	15.678								0
Nature Neuroscience	15.664								1
Journal of Experimental Medicine	15.612								0
Annals of Internal Medicine	15.516								0
Journal of Clinical Oncology	15.484								0
Nature Methods	15.478								6
Genes and Development	14.795								3
Nature Physics	14.677								0
PLoS Biology	13.501								2
Neuron	13.41								0
Molecular Cell	13.156								0
Circulation	12.755								0
PLoS Medicine	12.601								0
Developmental Cell	12.436								0
Gastroenterology	11.673								0
Genome Research	11.224								6
American Journal of Human Genetics	11.092								3
Nature Structural and Molecular Biology	11.085								0
Journal of the American College of Cardiology	11.054								0
Blood	10.896								0
Hepatology	10.734								0
Current Biology	10.539								0
Gut	10.015								0
British Medical Journal	9.723								0
Circulation Research	9.721								1
Plant Cell	9.653								0
Nano Letters	9.627								0
Journal of Cell Biology	9.598								0
PNAS	9.598								1
Molecular and Cellular Proteomics	9.425								7
PLoS Pathogens	9.336								0
American Journal of Psychiatry	9.127								0
American Journal of Respiratory and Critical Care Medicine	9.074								0
Annals of Neurology	8.813								0
PLoS Genetics	8.721								0

Repeatability

Ioannidis JPA, Allison DB, Ball CA, et al. Repeatability of published microarray gene expression analyses. *Nat Genet*. 2009; 41(2):149-155.

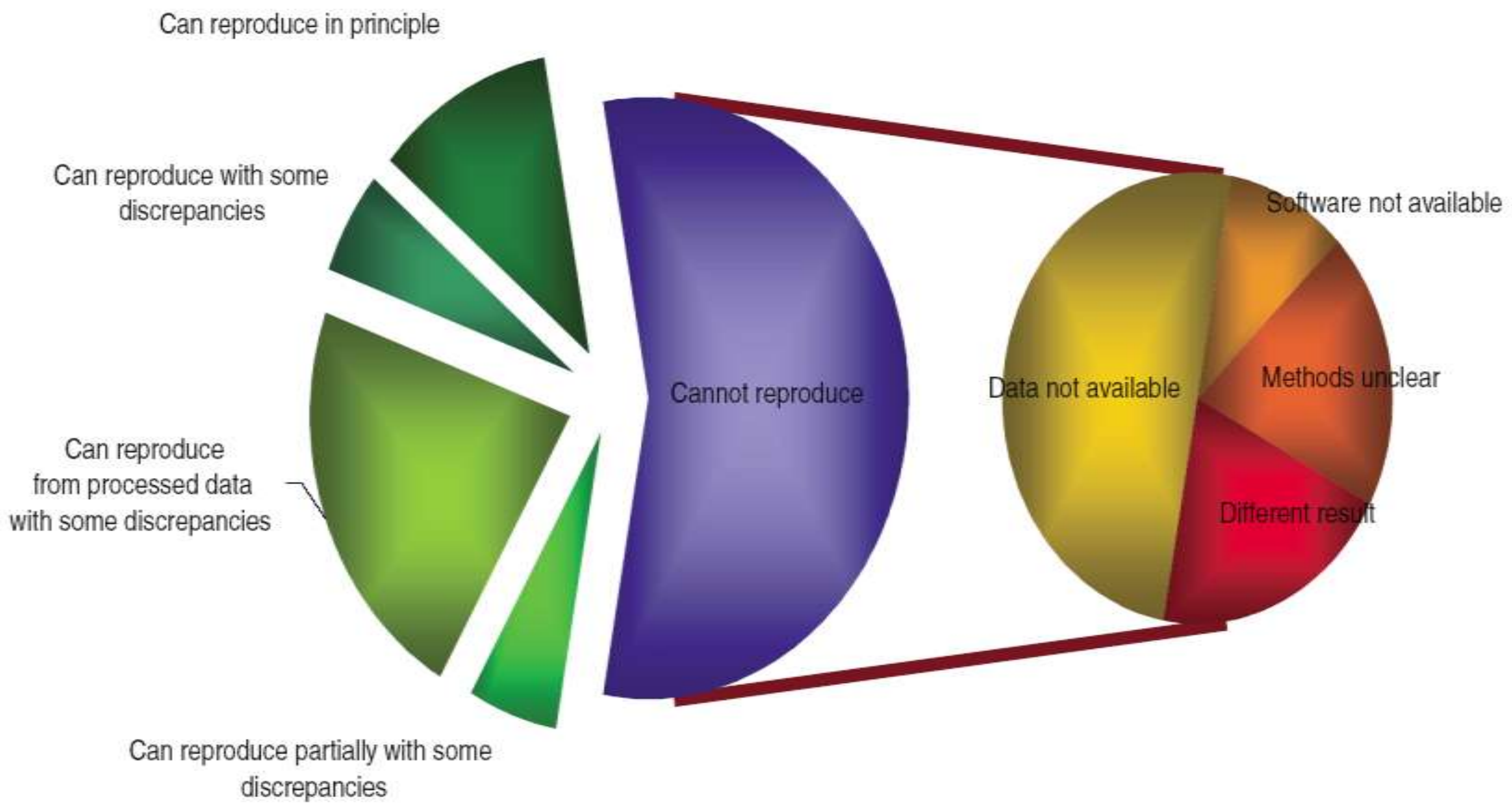


Figure 1 Summary of the efforts to replicate the published analyses.

Ioannidis JPA and Khoury MJ. Improving Validation Practices in "Omics" Research. *Science*. 2011; 334(6060):1230-1232.

Fig. 1. The validation of omics research for use in medicine and public health requires fulfilling multiple steps. [Adapted from (*7*)]

encemag.org **SCIENCE** VOL 000 MONTH 2011

From Ioannidis JPA and Khoury MJ. *Science*. 2011; 334(6060):1230-1232. Reprinted with permission from AAAS.

Table 4. Suggestions for the future of knowledge integration

Knowledge management	
	Methods for mining published and unpublished data
	Registration of observational datasets and, when appropriate, protocols
	Availability of raw data and analysis codes
	Facilitation of repeatability and reproducibility checks, replication culture
	Consideration of live stream information
Knowledge synthesis	
	Facilitation of consortia with prospective measurements
	Optimization of multiconsortial space, competition, and communication
	Prospective study networks
Knowledge translation	
	Anticipatory rather than *post hoc* brokering